

VU Research Portal

Returns to density in operations of the Netherlands Railways

Pels, E.; Daniel, V.E.; Rietveld, P.

2004

document version

Early version, also known as pre-print

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Pels, E., Daniel, V. E., & Rietveld, P. (2004). *Returns to density in operations of the Netherlands Railways*. (TI discussion paper; No. 2004-004/3). Tinbergen Instituut (TI).

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl



TI 2004-004/3

Tinbergen Institute Discussion Paper

Returns to Density in Operations of the Netherlands Railways

Eric Pels

Vanessa E. Daniel

*Piet Rietveld**

Department of Spatial Economics, Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam.

** Tinbergen Institute.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam

Roetersstraat 31

1018 WB Amsterdam

The Netherlands

Tel.: +31(0)20 551 3500

Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam

Burg. Oudlaan 50

3062 PA Rotterdam

The Netherlands

Tel.: +31(0)10 408 8900

Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to driessen@tinbergen.nl.

Most TI discussion papers can be downloaded at <http://www.tinbergen.nl>.

Returns to density in operations of the Netherlands Railways.

Eric Pels Vanessa E. Daniel Piet Rietveld

Free University, Department of Spatial Economics
De Boelelaan 1105, 1081 HV Amsterdam

abstract

Rail cost function analysis has been a popular topic in the (empirical) economics literature over the past decades. Most studies find increasing returns to density for rail companies. The results can, however, be quite diverse. Results for the Dutch National Railway company (NS), for instance, indicate very strong increasing returns to density (Andrikopoulos and Loizides, 1998) or decreasing returns to density (Preston, 1994). Using the, to our knowledge, most comprehensive dataset for the NS, this paper estimates a translog variable cost function for the NS. While the returns to density parameter shows increasing returns, like so many other papers, the calculated standard errors shows that the null hypothesis of constant returns may not be rejected.

1. Introduction

In discussions on a possible privatization of public railways, knowledge of returns to scale is important. When increasing returns to scale are present, straightforward marginal cost pricing leaves (private) railway operators with a loss. Public provision of rail services then requires a subsidy, or prices (in sub-markets) must be set above marginal costs. When returns to scale are present, privatization of a public railway may lead to the creation of a (local) natural monopoly, which cannot charge marginal costs without incurring losses. The results of both privatization of public railways and regulation of private railways are thus dependent on the presence of returns to scale.

Rail cost function analysis has been a popular topic in the (empirical) economics literature over the past decades. Wellington (1893) already discussed this issue (cited in Borts, 1952). Over the years the body of literature increased steadily, first using linear and Cobb-Douglas type cost functions or production functions, and later using flexible functional forms, while also efficiency analysis gained ground.

The abundance of literature allows for the comparison of the results of different studies. This may lead to confusing (even conflicting) views; for example, different studies conclude that the output elasticity of costs for the Dutch national

railway company is very low (so that the company operates under strong increasing returns to scale) or above one (implying decreasing returns to scale). Using time series to estimate a long run translog cost function with passenger kilometers and freight kilometers as outputs Van Ooststroom (1998) estimates an output elasticity of 0.52 for the Dutch railway company. Andrikopoulos and Loizides (1998) use cross section data and the sum of freight and passenger kilometers as outputs, and estimate an elasticity of 0.04 for the same company. Preston (1994) reports that the Dutch railway company is operating under decreasing returns to scale.

There may be various explanations for the difference in findings, such as the choice of the specific outputs, and the different types of data (limited time series versus pooled data). Moreover, although various studies report increasing or decreasing returns, one would typically need the associated standard error to come to a final conclusion.

In this paper, we calibrate a variable cost function of the Dutch Railways for the period 1951 to 1993, using, to our knowledge, the most extensive cost data for the Dutch Railways available. Based on these estimations, a measure of returns to scale (density) is calculated, together with the associated standard error.

The structure of the paper is as follows. Section 2 contains a concise review of the literature on railway cost function estimation. Section 3 discusses the methodology used in this paper, and Section 4 discusses the data. Section 5 contains the estimation results and Section 6 concludes.

2. Literature review

Rail cost function estimation has been a popular topic in the applied economics literature. Early examples include Borts (1960)¹, and usually rely on statistical cost-output relations or Cobb-Douglas cost functions. Statistical cost-output relations do not include input prices as explanatory variables. This implies that there may be a missing variable bias². Also, the Cobb-Douglas (and other inflexible functional forms) have their problems; such forms put a priori restrictions on the production technology, and may therefore be too restrictive.

¹ Earlier papers date back to the early 20th century or the late 19th century.

² When the firm does not have a costminimizing strategy but a cost function is estimated as if the firm does have such a strategy, the effects will be less severe; one simply includes too many variables.

Since the 1970s flexible functional forms, which do not put a priori restrictions on the production technology, gained popularity³. The most common form is the translog specification, which has been used in a number of rail cost function studies. Table 1 contains a number of cost function estimations for U.S. railroads. The reported numbers were obtained from underlying studies, or based on own calculations using inputs from underlying studies (see equation (6)); because the definition of economies of scale and density varies, some readjustments were necessary to make the numbers comparable. The general pattern shows constant to slightly increasing returns to scale and increasing returns to density⁴.

Table 1 about here

Friedlaender and Spady (1980) report some firms with decreasing returns to density. They attribute this to the fact that a short-run cost function is estimated. A firm facing short-run capital shortage may face short-run decreasing returns to density and long-run increasing returns to density. Friedlaender et al. (1993) report some very high values for the returns to density parameter. However, the associated standard errors are very large in these cases; generally speaking, returns to scale parameters with values larger than four appear to have very high standard errors. Smaller values for the returns to scale parameter are in many cases statistically different from 1. The rather high value for Brauetigam et al. (1982) is based on calculations using equation (6), and not reported by Brauetigam. The reported output elasticity of cost is, however, rather low (0.1105).

Table 2 about here

³ Caves and Christensen (1980) define the flexibility property for the case of consumer behavior as follows. "For any particular data point (one observation on prices and income) any set of price and income elasticities can be achieved through an appropriate set of parameter values. A nonflexible form is only capable of achieving a subset of the full range of price and income elasticities". In the case of cost (production) functions, this means that elasticities of factor substitution are unrestricted (and not necessarily constant); there are no a priori constraints on the first and second order derivatives of the cost function.

⁴ Generally speaking, returns to scale are defined as 1 over the sum of the output elasticity of cost and the network size elasticity of cost and returns to density are defined as 1 over the output elasticity of cost. Precise definitions will be provided later.

Table 2 briefly summarizes cost function estimations for European countries. At first sight, the variance in the returns to density parameter is much larger compared to U.S. studies. Preston (1994) reports a very high value for the Finnish railway company, and notes that this is a railway company with a relatively low traffic density per kilometer of line. All other values are, however, below the value of 6, which may still be a little high compared to the U.S. numbers. Also Andrikopoulos and Loizides (1998) present very high returns to density parameters. It is not entirely clear why these values are so high. It could be a result of the aggregate output measure (kilometers passengers plus kilometers freight); Filippini and Maggi (1992), however, use a similar output indicator and present returns to scale and density parameters that are of the same order of magnitude as the average values for the U.S. It could also be caused by the by the specification of the cost function itself; a time series estimation for specific European railway companies with only a few degrees of freedom. Other European studies present results that are similar to the U.S. results. McGeehan (1993) and van Ooststroom (1998) present values that still are a little high compared to most U.S. studies. They are, however, comparable to some of the values reported by Friedlaender et al. (1993).

Both studies of U.S. and European railways thus hint at constant returns to scale and increasing returns to density, although the variance in European studies tends to be larger. One explanation could be the different “orientation” of U.S. and European railway companies. Although European (and U.S.) companies may already be quite diverse between themselves, passenger transport is arguably more important for European companies. Moreover, institutional settings between European countries may simply be more heterogeneous. Looking at the various studies, we see that Caves et al. (1980), Caves et al. (1981) and Caves et al. (1981) use similar data and specifications. Also Berndt et al. (1993) and Friedlaender et al. (1993) use the same data and specifications. Other authors using U.S. data may use different data, but still use cross-sections or pooled data. Cantos (2000, 2001), using similar data in two studies (11 observations are added in one study) already finds markedly different returns to scale parameters. Andrikopoulos and Loizides (1998) have a similar data structure, although for different years, but with a different specification and different outputs they obtain completely different results. Other authors estimating cost functions for European railway companies use time series or pooled data.

Harris (1977) argues that “inappropriate measures of output” are frequently used, usually gross ton-miles. Although such an output provides a “common measure of the output of passengers and freight”, it also (wrongly) includes the weight of engines and cars, while freight and passengers carried are the true output. Moreover, using gross ton-miles one implicitly assumes that one ton carried 1000 miles and 1000 tons carried 1 mile are equivalent. Keeler argues that omitting the average length of haul in the specification leads to biased estimates⁵. Also, Keeler argues that there is a positive correlation between traffic density and the passenger/freight service ratio. Simply adding up passenger and freight gross ton-miles ignores this fact, and may lead to biased results. From Tables 1 and 2 it is clear that (many) different output specifications are used, and although this more than likely influences the results, it is impossible to say that “one performs better than the other”. Preston (1994) specifically concentrates on supply-oriented (intermediate) output measure, because final demand related output measures such as passengerkilometers and freight tonkilometers are influenced by e.g. fare restrictions and therefore “poor measures of managerial and organizational performance”. The “correct” output may thus be dependent on the structure, operating environment and legal status of the firm. In this paper, we follow Preston (1994) because the Dutch railway company is operating under strict fare controls.

3. Methodology

Translog.

The advantage in estimating a flexible functional form is that one puts no a-priori restrictions on cost elasticities (as is e.g. the case with the Cobb-Douglas function). Moreover, the translog may be used as a Taylor-series approximation of an unknown function at a certain point. Around that point, the first and second order derivatives of the translog function and the unknown function (and the dual functions⁶) are the same. The most simple translog specification for a micro-economic cost function is (subscripts denoting individual observations have been omitted to reduce notation.)⁷:

⁵ Keeler (1977) reports that the correlation between density and average length of haul is 0.77.

⁶ Blackorby and Diewert (1979)

⁷ The underlying assumption is that there exists a transformation process with a strictly convex input structure from which the maximum possible amount of outputs is produced. The dual to this transformation process is a multi-product cost function that minimises the cost of all inputs.

$$\begin{aligned}
\ln C = & \alpha_0 + \sum_{i=1}^m \alpha_i (\ln Y_i - \ln Y_i^*) + \sum_{j=1}^n \beta_j (\ln P_j - \ln P_j^*) + \phi (\ln K - \ln K^*) + \\
& \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \delta_{ij} (\ln Y_i - \ln Y_i^*) (\ln Y_j - \ln Y_j^*) + \\
& \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \gamma_{ij} (\ln P_i - \ln P_i^*) (\ln P_j - \ln P_j^*) + \frac{1}{2} \phi (\ln K - \ln K^*)^2 + \\
& \sum_{i=1}^m \sum_{j=1}^n \rho_{ij} (\ln Y_i - \ln Y_i^*) (\ln P_j - \ln P_j^*) + \\
& \sum_{i=1}^m \eta_i (\ln Y_i - \ln Y_i^*) (\ln K - \ln K^*) + \sum_{j=1}^n \lambda_j (\ln P_j - \ln P_j^*) (\ln K - \ln K^*) + \varepsilon
\end{aligned} \tag{1}$$

where C is operational cost, Y_i is output i and P_j is the price of input j . K is the quasi-fixed input, which is included because we estimate a (short-run) variable cost function (Caves et al., 1981). When the quasi-fixed factor would not be included in the estimations, the parameter estimates would be biased⁸. The quasi-fixed input is corrected for usage (Oum et al., 1991; Oum and Zhang, 1991).

The point of approximation is denoted by superscript $*$. When the translog is interpreted as a Taylor-series expansion, α_0 is the value of the unknown function at the point of approximation, and the parameters are the first and second-order derivatives of the unknown function with respect to the various variables. Symmetry implies that $\delta_{ij} = \delta_{ji}$ and $\gamma_{ij} = \gamma_{ji}$. Moreover, the restrictions for linear homogeneity in factor prices are:

$$\begin{aligned}
\sum_{j=1}^n \beta_j &= 1 \\
\sum_{i=1}^n \gamma_{ij} &= 0 \text{ with } j = 1, \dots, n \\
\sum_{j=1}^n \rho_{ij} &= 0 \text{ with } i = 1, \dots, m \\
\sum_{j=1}^n \lambda_j &= 0
\end{aligned}$$

Assuming cost minimisation and applying Shephard's lemma yields the following share equations, which are useful in the estimation procedure:

$$M_j = \beta_j + \sum_{i=1}^m \rho_{ij} (\ln Y_i - \ln Y_j^*) + \sum_{i=1}^n \gamma_{ij} (\ln P_i - \ln P_i^*) + \lambda_j (\ln K - \ln K^*) + \varepsilon_j; j = 1, \dots, n \quad (2)$$

Estimation.

The two standard methods used to estimate simultaneously all the equations of a fully specified structural model are three-stage least squares and full information maximum likelihood. The three-stage least squares estimation method combines the two-stage least squares and Zellner's method, the SUR (seemingly unrelated regressions, Zellner, 1962); this allows for correcting both simultaneous equation bias and disturbances' correlation between equations.

Zellner proposed a two step method to estimate models in which disturbances are contemporaneously correlated at a certain point of time. The initial difficulty when applying generalised least squares (GLS)⁹, is that the elements of the covariance matrix of the disturbance' terms are unknown. To alleviate this difficulty, Zellner proposed to apply OLS to each of the equations in a first step, and then to deduce the vector of the residuals e_i . The elements of the covariance matrix are computed as

$$s_{ii} = \frac{e_i' e_i}{n - k_i} \quad (3)$$

$$s_{ij} = \frac{e_i' e_j}{(n - k_i)^{1/2} (n - k_j)^{1/2}} \quad (4)$$

with n the number of equations, k_i the number of explanatory variables for each equation, and s_{ii} being the diagonal elements, s_{ij} being the other elements. The obtained estimation of the covariance matrix is then employed in the initial GLS methodology. It results in the feasible GLS estimator, or SUR estimator. Zellner's approach is the standard approach in the literature for the joint estimation of translog cost functions and share equations.

For the estimation process, it is important to realize that the point of approximation is the same for all observations. This means that the point of

⁸ When we derive a variable cost function, there will be a fixed factor as an argument in the variable cost function. When this factor is not included, the parameters would suffer from a "missing variable-bias".

⁹ The GLS is to be applied when disturbances are non spherical: $E(uu') = \sigma^2 \Omega$; it consists in the application of OLS including the covariance matrix in the specification of the estimator

standardization is no more than an adjustment of some of the parameters. Specifically, we can rewrite the share equations as:

$$M_j = \left(\beta_j - \sum_{i=1}^m \rho_{ij} \ln Y_j^* - \sum_{i=1}^n \gamma_{ij} \ln P_i^* - \lambda_j \ln K^* \right) + \sum_{i=1}^m \rho_{ij} \ln Y_i + \sum_{i=1}^n \gamma_{ij} \ln P_i + \lambda_j \ln K + \varepsilon; j = 1, \dots, n \quad (5)$$

The implication is that the constants in the share equations are not equal to β_j unless $Y_i^* = P_j^* = K_j^* = 1$. At any other point, the constants are estimated as the first RHS-terms in brackets in equations (5). β_j can be calculated from these constants. This means that one should calculate the proper β_j for each iteration in the iterative estimation procedure¹⁰. In this paper, we use full information maximum likelihood, where this is not an issue. See Appendix A for a similar discussion on the cost function.

Returns to scale and returns to density.

Harris (1977), points out that economies of scale are a property of a long-run cost function; average costs decrease as the firm size increases, where firm size is measured in outputs and network size. For (short-run) pricing issues (with fixed networks) economies of density (keeping network size fixed) are arguably more important (Harris, 1977).

It is, however, not uncommon to determine both returns to density and returns to scale from a short run variable cost function; for example, Caves et al. (1981) specify a short-run variable cost function and calculate returns to scale using two measures including and excluding network characteristics. Following Panzar and Willig (1977), it is common to define returns to scale as 1 over the sum of the output elasticities, where the output elasticity is $\partial \ln C / \partial \ln Y_i$ ¹¹. When the expression is larger than 1, positive returns to scale prevail¹². In the case of a short-run cost

¹⁰ Unless the data are standardized on the point $\ln(1)=0$.

¹¹ Christensen and Greene (1976) calculate the returns to scale measure as $1 - \sum_j \partial \ln C / \partial \ln Y_j$,

which is positive when scale economies prevail and negative when diseconomies prevail. This definition is also used in the transport economics literature for returns to density. It is straightforward to “translate” a number calculated from one definition to a number calculated from the other definition.

¹² From the dual production function we could determine an elasticity of scale. This elasticity of scale only equals the inverse of the output elasticity of the cost function when technology is homothetic.

function, (short-run) returns to scale are (1 minus the elasticity of cost with respect to capital) over the sum of the output elasticities.

In the transport economics literature, this definition is commonly used for (short run) returns to density (*RTD*); see e.g. Caves et al. (1981):

$$\begin{aligned}
 RTD &= \frac{1 - \frac{\partial \ln C}{\partial \ln K}}{\sum_j \frac{\partial \ln C}{\partial \ln Y_j}} \\
 &= \frac{1 - \left[\phi + \varphi (\ln K - \ln K^*) + \sum_{i=1}^m \eta_i (\ln Y_i - \ln Y_i^*) \right]}{\sum_i \left(\alpha_i + \frac{1}{2} \sum_{j=1}^m \delta_{ij} (\ln Y_j - \ln Y_j^*) + \eta_j (\ln K - \ln K^*) \right)} \quad (6)
 \end{aligned}$$

Returns to scale are usually calculated as $1/(\text{sum of output elasticities plus the network size elasticity})$. Available miles of track are frequently used in the rail cost literature as a measure of network size, see e.g. Bitzan (2000), Cantos (2001) and Caves et al. (1984). The theoretical interpretation of *RTS* may, however, be somewhat tedious when the network size variable cannot be interpreted as an output. When the outputs are passenger kilometers or revenue kilometers, network size (network length) could be seen as an input. We then calculate *RTS* using the elasticity of costs with respect to both outputs and an input.

Preston (1994) argues that many rail companies operate(d) in a regulated environment, and have little control over the final (consumer) prices. Preston therefore uses “length of line” as an “intermediate” output, so that the network size variable has an interpretation as an output and the calculation of *RTS* is solely based on outputs. Although the Netherlands Railways company also operates under strict fare controls, so that a measure of network size could be included as an (intermediate) output, in our case it was not possible to calibrate a meaningful cost function including a measure of network size due to data limitations. We therefore do not calculate *RTS*.

In Appendix 1 we show that it does not matter for the calculation of *RTD* (or *RTS*) which point of approximation is chosen. Estimating one cost function and

When technology is homothetic, positive returns to scale will still prevail when the output elasticity of

calculating *RTD* for all sample points yields the same results as calculating *RTD* from calibrated cost functions standardized at each sample point.

4. Data.

In this section, the data are described. We start with a concise description of the Netherlands Railway Company.

The Netherlands Railways: Nederlandse Spoorwegen (NS)

EU directive 91/440/EEC, implemented January 1 1993, recommended financial equilibrium, autonomy and competitiveness for European railways, and stated that transport services and infrastructure exploitation have to be discriminated. Indeed, in 1991, Netherlands Railways (NS) has split in five sub-sections: NS travellers, NS cargo, NS stations, NS real estate, and NS traffic control (Van Ooststroom, 1999). Data after 1993 are different from those before 1993. To keep homogeneity in the time series used in the analysis, the dataset is therefore not expanded beyond 1993.

The State, exclusive owner of NS, has gradually stopped any financial support¹³, and NS is now financially independent. Actually, NS' investments have been reduced over the years; the current value of capital decreased at a very slow pace over the four considered decades, with a large break between 1971 and 1972 (-8.5%), when part of the network was put out of operation.

Supply and demand

The supply of passenger transport services has almost tripled in the past forty years, when both the number of available seats and the covered distances are considered, see Figure 1.

From the 1970s, NS made increased the frequencies of trains (van Ooststroom, 1998), which becomes apparent in Figure 1; there is a sharp increase in the number of seat-kilometers. This increase is only temporary, and slows down following the oil crisis. In the 1980s, there was a growth again, with a rapid increase in the early 1990s.

cost is smaller than 1.

¹³ The European Industrial Relations Observatory, 'Comparative study on industrial relations in the rail transport sector, the Netherlands', and 'Les relations industrielles dans le secteur ferroviaire' <http://www.eiro.eurofound.ie/2000/03/word/nl000176s.doc>
<http://www.eiro.eurofound.ie/2000/03/Study/TN0003277S.html>

Figure 1 about here

Figure 1 shows that the actual demand for transport passenger services (traveller-kilometres) has been increasing at a relatively slow pace until the late 1980s, then growing sharply and stabilising around 1991, whereas the supply continued to increase. The sudden increase in demand in the early 1990s was the consequence of the government's decision to give students free access to public transport. As a matter of fact, the load factor, which was not far from fifty percent in the early 1950s, has diminished approximately by ten percentage points (Figure 2). Again, free student access to public transport is visible in 1991 with a sudden rise in the load factor. The load factor does, however, not reach the level of the early 1950s.

Figure 2 about here

Network

Until the early 1970s, the size of the network remained steady. After the early 1970s, the number of stations increased little by little, while at the same time the length of the network decreased, in a smaller scale, however. Unprofitable lines were deleted, and more stations were opened along “thick” lines. NS has strongly focused its attention on passenger transport. Freight operations are limited.

Figure 3 about here

The dataset.

The dataset contains a time series for the Dutch railways for the period 1951-1993, and was extracted for a great part from Van Ooststroom (1999). These data were combined with data from the online version of CBS data base, and from the yearly paper versions of the CBS review *Nederlandse spoorwegen, infrastructuur en rollend materiaal*, 31 december, in *Statistiek van het personenvervoer*.

In the short-run variable cost function, we include the following variables. The *operating costs* or *variable costs* used in the analysis break down into three components: labour costs, energy costs and maintenance costs (Figure 4). Input prices (Figure 5) are computed as the ratios between expenditures and corresponding volumes. For the price of labour, personnel expenditures are divided by the number of

employed persons. The energy cost is the ratio of energy expenditures for traction to the need for energy in kilowatt-hour. Finally, maintenance costs are obtained by dividing maintenance and materials expenditures by the sum of the number of train-kilometres goods and train-kilometres passengers. The greater part of the operating costs of the Dutch railways are labour costs (reaching 80% of total operating costs in the late 1970s). Labour became more and more expensive until the 1970s. The cost shares have been computed from the sum of these three items. Costs are deflated by dividing its initial value by the consumer price index, base 100 in 1980.

Figure 4 about here

The number of *seat-kilometres* is the only *output* used in the cost function. The NS gradually reduced the network length solely devoted to freight transport. Moreover, the number of freight-ton kilometers is small compared to the number of passenger-kilometers. Including freight as an output results in a theoretically incorrect sign for the first order parameter. One output (passenger-kilometers) is therefore used in the reported estimations. The number of stations was also considered as an (intermediate) output, but also has an incorrect sign. The number of stations may also be used as a measure of network size (but then one also would not expect a negative sign) or as a fixed factor. But in the latter case, one would expect the number of stations to be included in the value of capital, so that it is counted double. It was therefore dropped from the estimations all together.

Figure 5 about here

The *value of capital* is used as a quasi-fixed factor, and is simply the value of capital corrected for usage by applying the load factor. The load factor is the ratio of passenger-kilometres to capacity seat-kilometres. The value of capital is given in constant guilders, and is highly correlated with the total network length in kilometers.

Figure 6 about here

Finally, a *quadratic* trend was included to capture temporal effects.

5. Estimation results.

The estimation results are presented in Table 3. The Durbin-Watson statistic for the cost function is in the “safe-zone”. The R^2 is very high, which is due to the relatively large number of second-order terms.

All the squared terms have positive signs, which is according to expectations: the cost function is globally increasing in input prices and outputs, all other variables held constant. From the sign of the interaction terms, it appears that labour and energy are substitutes. This may be explained by the fact that the same number of seat-kilometers can be obtained by lengthening trains, keeping all other variables fixed, less drivers and more energy is needed, although the amount of energy needed is probably not doubled. Moreover, labour and maintenance appear to be substitutable goods.

The trend variable has a significant second-order coefficient with a negative sign. The constant represents the cost function value at the point of approximation, so that the negative sign of the second order term indicates that the cost evaluated at a given point (the point of approximation) decreases over time: productivity increases. Note that the estimated constant changes when we change the point of approximation. The interpretation of the second-order trend term does, however, not change. As outlined in the Appendix, the point of approximation is irrelevant in the discussion of *RTD* which will follow (see Appendix 1 for details).

Table 4 and Figure 7 present the results on the returns-to-density computed from the estimation results in table 3.

Tables 3 and 4 about here

Figure 7 about here

Up until 1967 the *RTD*-coefficient is fairly stable around 1; slight decreasing returns to density prevail. After 1968 increasing returns to density prevail. This is consistent with the above mentioned strategy change in the early 1970s (discussed by van Ooststroom, 1998). A demand oriented strategy was replaced by a supply oriented strategy. The frequency of service was increased, also off-peak, hoping that demand would pick up. As a result, the number of available seatkilometers increased. Because this was a “nation-wide” policy, more frequent trains automatically meant shorter

trains (given the number of carriages). The results show that lower cost per seat-kilometer could be obtained by increasing the number of seats (by making trains longer or introducing different train types). This is what we see in the late 1980s, when double-decker trains were introduced. The “old” carriages (replaced by the double-deckers) were used to increase the length of other trains, so that the total number of available seats increased: there are less opportunities to exploit density economies.

A general similar pattern (increasing trend) in *RTD* is found by van Ooststroom (1998), although in van Ooststroom’s case there are more fluctuations around the general trend. Our study thus confirms this increasing trend. However, judging by the standard error reported in Table 4, the value 1 falls inside the 95% confidence interval for *RTD* for each year, so that the null hypothesis of constant returns to scale cannot be rejected in each year¹⁴.

Very strong increasing returns to density, as reported by Andrikopoulos and Loizides (1998) for the NS, are not encountered in this study. Likewise, Preston (1994) reports that *RTD* for the NS is 0.77 in 1990. Although we find values for *RTD* in that order of magnitude, we certainly do not find it for the year 1990. The different results may be contributed to the set-up of the different studies. Andrikopoulos and Loizides (1998) use a different output and have a relatively small number of degrees of freedom. Preston (1994) uses a pool of different companies, so that the performance of the NS is judged against the best practice in Europe, rather than evaluated over time, as is done in this study.

6. Conclusion

Preston (1994) discusses the “traditional” and “revisionist” view of railway economists on the issue of the effect of firm size on costs. In short, the traditional view is that “size matters”; the railway industry faces declining average costs. As a result, most railway companies in Western Europe at some point in time became (state) monopolies, because it is cheaper to integrate smaller companies. According to the revisionist view, “size does not matter”, at least not in railway operations. Moreover, railway operations are thought to be contestable, and large state-owned

¹⁴ Standard errors are approximated using the delta-method (see e.g. Greene, 2000).

railway companies experience X-inefficiency. The revisionist view suggests that infrastructure and railway operations should be separated, and that competition between different railway operators is advisable; since the early 1990s, this has been the U.K. rail policy.

Although the estimated patterns in returns to density more or less reflect the findings of van Ooststroom (1998) for the Netherlands national railway company (NS), the associated standard errors univocally lead to the conclusion that constant returns to scale prevail through the entire sample period. In this sense, the results support (part) of the revisionist view for the Netherlands. Estimates of the returns to density parameter yield values that are larger than 1 (as in van Ooststroom (1998)) and smaller than 1 (as reported by Preston (1994) for the NS), but always statistically equal to 1. Very high values, such as in Andrikopoulos and Loizides (1998) were not encountered.

These findings suggest that it was a good point to restructure the Dutch rail sector by separating infrastructure and operations, and allowing competition. There are, however, a few comments. Firstly, X-inefficiency and the relation between the infrastructure operator and the rail operator(s) are ignored in this paper. Some of the problems faced by the Dutch railway company today are due to capacity problems of the infrastructure operator. Increasing competition “on the tracks” may be a good idea to increase the efficiency of rail operators. But when infrastructure capacity is limited, the railway companies may not be able to operate their optimal schedules, so that the (welfare economic) effect of competition is reduced. Secondly, this argument ignores network externalities on the demand side. Passengers would benefit from a merger between complementary local rail operators. When these rail operators are independent, (some) passengers have to switch operators to travel between the origin and the destination. A price increase by one operator then automatically decreases demand for the complementary operator. When the operators merge, they internalize this effect. This is, however, a demand effect, and conclusions cannot be drawn based upon the results (on cost effects) in this paper.

References

Blackorby, C. and W.E. Diewert (1979), Expenditure functions, local duality and second order approximations, *Econometrica*, **47**(3), 579-601

- Andrikopoulos, A.A. and J. Loizides (1998), Cost structure and productivity growth in European railway systems, *Applied Economics*, **30**, 1625-1639
- Berndt, E.R. et al. (1993), Costs effects of mergers and deregulation in the U.S. rail industry, *Journal of Productivity Analysis*, **4**, 127-144
- Bitzan, J. (2000), *Railroad cost conditions – implications for policy*, Report prepared for the Federal Railroad Administration, U.S. Department of Transportation.
- Borts, G.H. (1952), Production relations in the railway industry, *Econometrica*, **20**, 71-79
- Braeutigam, R.R., A.F. Daughety and M.A. Turnquist (1984), A firm specific analysis of economies of density in the U.S. railroad industry, *Journal of Industrial Economy*, **33**, 3-20
- Braeutigam, R.R., A.F. Daughety and M.A. Turnquist (1982), The estimation of a hybrid cost function for a railroad firm, *Review of Economics and Statistics*, **64**, 394-404
- Brown, R.S., D.W. Caves and L.R. Christensen (1979), Modelling the structure of cost and production for multiproduct firms, *Southern Economic Journal*, **46**, 256-273
- Caves, D.W., and L.R. Christensen (1980), The relative efficiency of public and private firms in a competitive environment: the case of Canadian railroads, *Journal of Political Economy*, **88**, 958-976
- Caves, D.W., L.R. Christensen and J.A. Swanson (1981), Economic performance in regulated and unregulated environments: a comparison of U.S. and Canadian railroads, *Quarterly Journal of Economics*, **96**, 559-581
- Caves, D.W., L.R. Christensen and J.A. Swanson (1981), Productivity growth, scale economies and capacity utilization in U.S. railroads, 1955-1974, *American Economic Review*, **71**, 994-1002
- Cantos, P. (2001), Vertical relationships for the European railway industry, *Transport Policy*, **8**, 77-83
- Cantos, P. (2000), A subbadditivity test for the cost function of the principal European railways, *Transport Reviews*, **20**, 275-290
- Cantos, P. and J. Maudos (2000), Technical change in the European rail sector: a stochastic frontier approach, *International Journal of Transport Economics*, **27**, 55-76
- Christensen, L.R. and W.H. Greene (1976), Economies of scale in U.S. electric power generation, *Journal of Political Economy*, **84**(4), 655-676
- De Borger, B. (1993), The economic environment and public enterprise behaviour: Belgian railroads, 1950-1986, *Economica*, **60**, 443-463
- De Borger, B. (1992), Estimating a multiple-output generalized Box-Cox cost function: Cost structure and productivity growth in Belgian railroad operations, 1950-1986, *European Economic Review*, **36**, 1379-1398

- De Borger, B. (1991), Hedonic versus homogeneous output specifications of railroad technology: Belgian railroads 1950-1986, *Transportation Research*, **36A**, 633-644
- Filippini, M. and R. Maggi (1992), The cost structure of Swiss private railways, *International Journal of Transport Economics*, **19**(3), 307-327
- Friedlaender, A.F., E.R. Berndt, J.S.-E. W. Chiang, M. Showalter and C.A. Velluro (1993), Rail costs and capital adjustments in a quasi-regulated environment, *Journal of Transport Economics and Policy*, **27**, 131-152
- Greene, W.H. (2000), *Econometric analysis*, Prentice-Hall
- Griliches, Z. (1972), Cost allocation in railroad regulation, *Bell Journal of Economics and Management Science*, **3**, 26-41
- Harris, R.G. (1983), Economies of traffic density in the rail freight industry, *Bell Journal of Economics*, **8**, 556-564
- Harris, R.G. and C. Winston (1983), Potential benefits of rail mergers: an econometric analysis of network effects on service quality, *Review of Economics and Statistics*, **65**, 32-40
- Hasenkamp, G. (1976), A study of multiple-output production functions: Klein's railroad study revisited, *Journal of Econometrics*, **4**, 253-262
- Ivaldi, M., G.J. McCullough (2001), Density and Integration effects on Class I U.S. freight railroads, *Journal of Regulatory Economics*, **19**, 161-182
- Keeler, T.E. (1974), Railroad costs, returns to scale and excess capacity, *Review of economics and statistics*, **56**, 201-208
- Kim, H.Y. (1987), Economies of scale and scope in multiproduct firms: evidence from US railroads, *Applied Economics*, **19**, 733-741
- Loizides, J. and E.F. Tsionas (2002), Productivity growth in European railways: a new approach, *Transportation Research*, **36A**, 633-644
- McGeehan, H. (1993), Railway costs and productivity growth: The case of the republic of Ireland, 1973-1983, *Journal of Transport Economics and Policy*, **27**, 19-32
- Ooststroom, H.v. (1998), *Marktwerving en regulering bij spoorwegen: Theorie, emoirie en beleid*, Ph.D. dissertation, Free University, Amsterdam
- Preston, J. (1994), *Does size matter? A case study of Western European Railways*, unpublished paper, Institute of Transport Studies, University of Leeds
- Wilson, W.L. (1997), Cost savings and productivity in the railroad industry, *Journal of Regulatory Economics*, **11**, 21-40

Appendix 1 Choice of approximation point

To show the effect of the point of approximation on the parameter estimates and the estimate of RTS , we consider, without loss of generality, the following simplified cost function:

$$\ln C = \alpha_0 + \sum_{i=1}^m \alpha_i (\ln Y_i - \ln Y_i^*) + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \delta_{ij} (\ln Y_i - \ln Y_i^*) (\ln Y_j - \ln Y_j^*) + \varepsilon \quad (A1)$$

Calculating $\sum_{i=1}^m \frac{\partial \ln C}{\partial \ln Y_i}$ from (A1) yields (with $m=2$ and $\delta_{12}=\delta_{21}$)

$$\alpha_1 + \alpha_2 + \delta_{1,1} (\ln Y_1 - \ln Y_1^*) + \delta_{1,2} (\ln Y_1 + \ln Y_2 - \ln Y_1^* - \ln Y_2^*) + \delta_{2,2} (\ln Y_2 - \ln Y_2^*) \quad (A2)$$

In the estimation of (A1), it is important to realize that the point of approximation, $\ln Y^*$, is a constant; it is the same for all parameters. This has implications for the parameter estimates. For $m=2$, we can rewrite the cost function as:

$$\begin{aligned} \ln C = & \left(\alpha_0 - \alpha_1 \ln Y_1^* - \alpha_2 \ln Y_2^* + \frac{1}{2} \delta_{1,1} (\ln Y_1^*)^2 + \frac{1}{2} \delta_{2,2} (\ln Y_2^*)^2 + \delta_{1,2} \ln Y_1^* \ln Y_2^* \right) + \\ & \alpha_1 \ln Y_1 + \alpha_2 \ln Y_2 + \frac{1}{2} \delta_{1,1} (\ln Y_1)^2 + \frac{1}{2} \delta_{2,2} (\ln Y_2)^2 + \delta_{1,2} \ln Y_1 \ln Y_2 - \\ & \delta_{1,1} \ln Y_1^* \ln Y_1 - \delta_{1,2} \ln Y_1^* \ln Y_2 - \delta_{1,2} \ln Y_2^* \ln Y_1 - \delta_{2,2} \ln Y_2^* \ln Y_2 + \varepsilon \Leftrightarrow \end{aligned} \quad (A3)$$

$$\ln C = \varpi_0 + \varpi_1 \ln Y_1 + \varpi_2 \ln Y_2 + \frac{1}{2} \delta_{1,1} (\ln Y_1)^2 + \frac{1}{2} \delta_{2,2} (\ln Y_2)^2 + \delta_{1,2} \ln Y_1 \ln Y_2 + \varepsilon \quad (A4)$$

where

$$\begin{aligned} \varpi_0 &= \alpha_0 - \alpha_1 \ln Y_1^* - \alpha_2 \ln Y_2^* + \frac{1}{2} \delta_{1,1} (\ln Y_1^*)^2 + \frac{1}{2} \delta_{2,2} (\ln Y_2^*)^2 + \delta_{1,2} \ln Y_1^* \ln Y_2^* \\ \varpi_1 &= \alpha_1 - \delta_{1,1} \ln Y_1^* - \delta_{1,2} \ln Y_2^* \\ \varpi_2 &= \alpha_2 - \delta_{1,2} \ln Y_1^* - \delta_{2,2} \ln Y_2^* \end{aligned}$$

so that

$$\begin{aligned}\hat{\alpha}_0 &= \hat{\omega}_0 + \hat{\alpha}_1 \ln Y_1^* + \hat{\alpha}_2 \ln Y_2^* - \frac{1}{2} \hat{\delta}_{1,1} (\ln Y_1^*)^2 - \frac{1}{2} \hat{\delta}_{2,2} (\ln Y_2^*)^2 - \hat{\delta}_{1,2} \ln Y_1^* \ln Y_2^* \\ \hat{\alpha}_1 &= \hat{\omega}_1 + \hat{\delta}_{1,1} \ln Y_1^* + \hat{\delta}_{1,2} \ln Y_2^* \\ \hat{\alpha}_2 &= \hat{\omega}_2 + \hat{\delta}_{1,2} \ln Y_1^* + \hat{\delta}_{2,2} \ln Y_2^*\end{aligned}$$

The $\hat{\omega}$'s are estimated from the data. The $\hat{\alpha}$'s can then be calculated from the estimated parameters. When the point of approximation is changed, $\hat{\delta}$ and $\hat{\omega}$ will not change (because $\ln Y^*$ is not an explanatory variable in A4). But $\hat{\alpha}$ will change: changing the point of approximation breaks down to changing the first-order parameters (and thus the estimate of *RTD* at the point of approximation).

For the calculation of *RTD* for *specific* points, the point of approximation does not matter. It follows from A4 that

$$\sum_i \frac{\partial \ln C}{\partial \ln Y_i} = \hat{\omega}_1 + \hat{\omega}_2 + \hat{\delta}_{1,1} (\ln Y_1) + \hat{\delta}_{1,2} (\ln Y_1 + \ln Y_2) + \hat{\delta}_{2,2} (\ln Y_2) \quad (\text{A5})$$

which is invariant to the point of approximation. When one calculates $\partial \ln C / \partial \ln Y_i$ directly from (A1), then one obtains

$$\sum_i \left. \frac{\partial \ln C}{\partial \ln Y_i} \right|_{Y_i=Y_i^*} = \alpha_1 + \alpha_2 \quad (\text{A6})$$

It is thus important to realize that $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are not estimated directly, but calculated from $\hat{\omega}_i$. Moreover, substituting the expression for $\hat{\omega}_i$ in (A6) yields (A5), and since $\hat{\omega}_i$ does not matter on the point of approximation, we have shown that the estimate of *RTD* for specific points is invariant to the point of approximation. The standard errors will also be unaffected, because the standardization has no effect on the variance of the explanatory variables.

	RTD	RTS	period	functional form	horizon	outputs	obs
Berndt et al. (1993)	1.56		1974-1986	translog	short	tonmiles freight	229
Bitzan (2000)	1.87-1.91	0.98-1.21	1983-1997	translog	long	tonmiles freight	215
Brauetigam et al. (1984)	2.71, 3.21		1976-1978	translog,	long, short	monthly carloads speed	35
Brauetigam et al. (1982)	19.02		1969-1977	translog,	short	loaded car miles, speed	108
Brown et al. (1979)	1.193		1936	translog	long	tonmiles freight, passenger miles	67
Caves et al. (1980)	1.13-1.18		1955, 1963, 1974	translog,	long	tonmiles freight, passengermiles	154
Caves et al. (1981)	1.01-1.23		1955, 1963, 1974	translog	short, long	tonmiles freight, passengermiles, average length of freight haul	154
Caves et al. (1981) (includes Canada)	1.02-1.26		1955, 1963, 1974	translog,	short	tonmiles freight, passengermiles, average length of freight haul	160
Friedlaender and Spady (1980)	0.72-1.63		1968-1970	translog,	short	revenue tonmiles	75
Friedlaender et al. (1993)	1.41-30.3	0.93-2.69	1974-1986	translog	short, long	tonmiles of freight	229
Hasenkamp (1976)		1.13-1.16	1928, 1936	Cobb-Douglas, CES	long	tonmiles freight, passengermiles	86, 71
Ivaldi and McCullough (2001)	1.9	1.65	1978-1997	translog	short	carmiles bulk traffic, carmiles high-value traffic, carmiles general traffic	299
Wilson (1997)	1.11-1.60	0.86-1.05	1978-1989	translog	short	revenue tonmiles	304

Table 1 Cost function estimations for U.S. railroads

	RTD	RTS	period	functional form	horizon	outputs	obs
Cantos (2001) Europe	1.42-2.04	0.47-2.06	1973-1990	translog	long	trainkilometers freight, passenger trainkilometers	204
Cantos (2000) Europe	1.90-2.34	0.25-1.48	1973-1990	translog	long	trainkilometers freight, passenger trainkilometers	215
Cantos and Maudos (2000)	0.77-1.81		1970-1990	translog (frontier)	long	trainkilometers freight, passenger trainkilometers	306
de Borger (1993) Belgium	0.85-1.34		1950-1986	translog,	short	hedonic passenger kilometers, hedonic ton kilometers	37
de Borger (1992) Belgium	0.85-1.44		1950-1986	generalized Leontieff, generalized Box-Cox, translog	short	hedonic passenger kilometers, hedonic ton kilometers	37
de Borger (1991) Belgium	0.62-1.65		1950-1986	translog,	short	passenger kilometers, ton kilometers	37
Andrikopoulos and Loizides (1998) Europe	5.54-47.2		1969-1993	translog	long	traffic units (kilometer passengers + kilometers freight)	25
Filippini and Maggi (1992) Switzerland	1.45-2.11	0.95-1.62	1985-1988	translog,		aggregate wagon kilometers for passenger and freight	192
McGeehan (1993) Ireland	4.07		1973-1983	translog,	short	freight tonmiles passenger miles	44
van Ooststroom (1998) Netherlands	1.09-4.17		1951-1991	translog	long	trainkilometers of pax trainkilometers of freight	40
Preston (1994) Europe	0.77-53.8	0.60-1.26	1971-1990	translog	long	trainkilometers, length of line	300

Table 2 Cost function estimations for European railroad companies

Explanatory variable	Parameter estimate	Standard error
constant	21.246	0.045
output in seat-km	0.504	0.082
labour price	0.684	0.008
energy price	0.093	0.007
maintenance price	0.223	0.001
value of capital corrected for usage	0.229	0.103
(output) ² /2	0.057	0.446
(labour price) ² /2	0.127	0.015
(energy price) ² /2	0.035	0.012
(maintenance price) ² /2	0.085	0.014
(value of capital corrected for usage) ² /2	2.677	1.344
output × price of labour	-0.048	0.027
output × price of energy	0.023	0.029
output × price of maintenance	0.025	0.012
labour price × energy price	-0.039	0.011
labour price × maintenance price	-0.088	0.007
energy price × maintenance price	0.003	0.008
value of capital corrected for usage × output	0.598	0.586
value of capital corrected for usage × labour price	0.004	0.092
value of capital corrected for usage × energy price	0.050	0.083
value of capital corrected for usage × maintenance price	-0.054	0.023
trend	0.001	0.003
trend ²	-0.0002	0.0000
DW (cost function)	1.48	
adjusted R ² (cost function)	0.99	
observations	43	

Table 3 Estimation results of railway cost function, FIML, specification (1) including trend

Year	RTD	Standard error
1951	0.625	0.525
1952	0.878	0.583
1953	0.976	0.596
1954	0.820	0.488
1955	0.698	0.437
1956	0.725	0.422
1957	0.947	0.438
1958	1.012	0.447
1959	1.043	0.448
1960	0.956	0.390
1961	1.013	0.366
1962	0.748	0.380
1963	0.604	0.401
1964	0.712	0.384
1965	0.924	0.362
1966	1.057	0.366
1967	1.189	0.397
1968	1.474	0.499
1969	1.817	0.604
1970	2.033	0.594
1971	1.998	0.547
1972	3.179	1.629
1973	2.699	1.162
1974	2.445	0.888
1975	2.637	1.101
1976	3.044	1.579
1977	3.102	1.682
1978	2.792	1.342
1979	2.169	0.791
1980	2.328	0.850
1981	2.122	0.671
1982	1.717	0.375
1983	2.542	1.014
1984	2.804	1.267
1985	2.689	1.144
1986	3.217	1.617
1987	3.274	1.554
1988	3.160	1.384
1989	3.076	1.233
1990	2.812	0.938
1991	0.851	0.640
1992	1.168	0.646
1993	1.472	0.640
sample mean	1.528	0.377
1951-1960	0.863	0.470
1961-1970	1.112	0.365
1971-1980	2.618	1.107
1981-1990	2.702	1.035

Table 4 Returns to density in Netherlands Railways

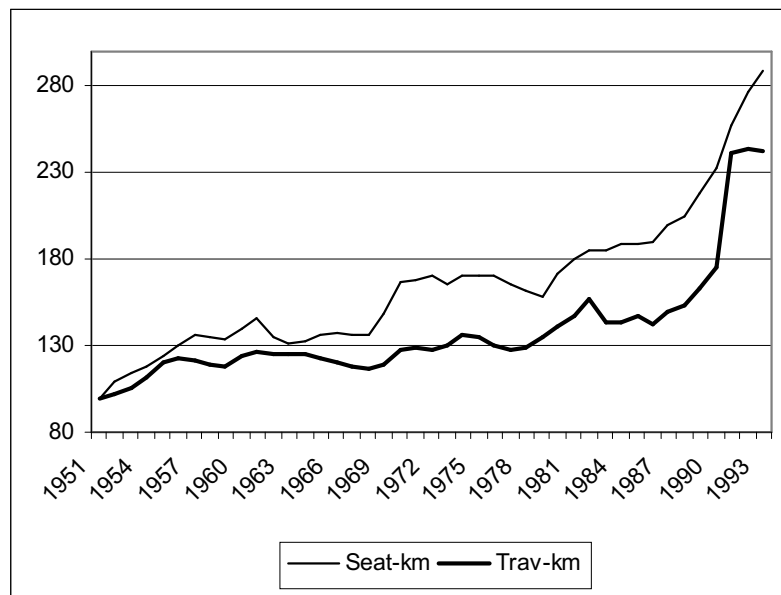


Figure 1 Evolution of seat-km and traveller-km; NS, base 100=1951

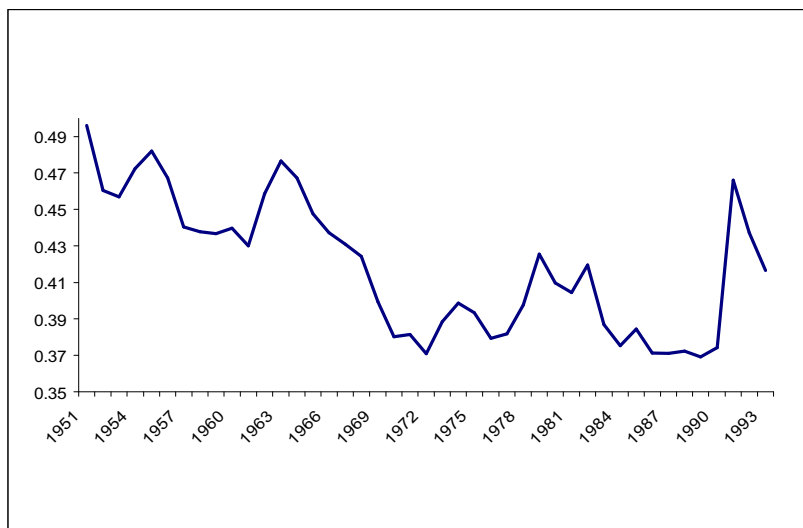


Figure 2 Development of load factor over time, NS.

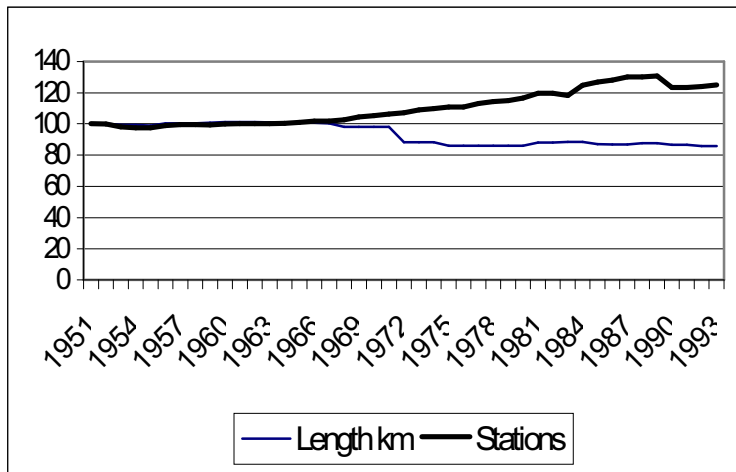


Figure 3 Evolution of the size of the network; NS, base 100=1951

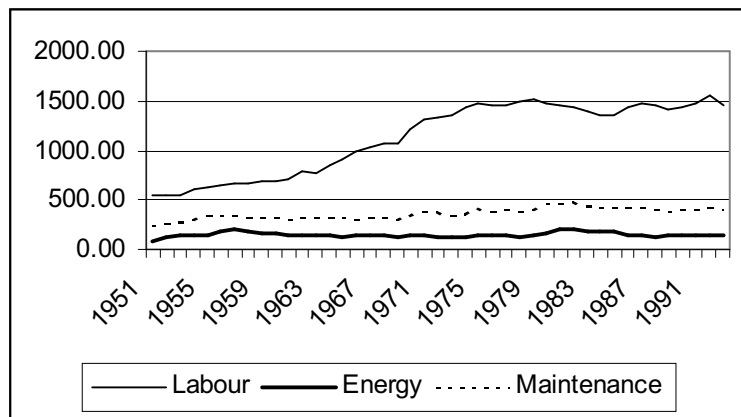


Figure 4 Operating costs over time, NS

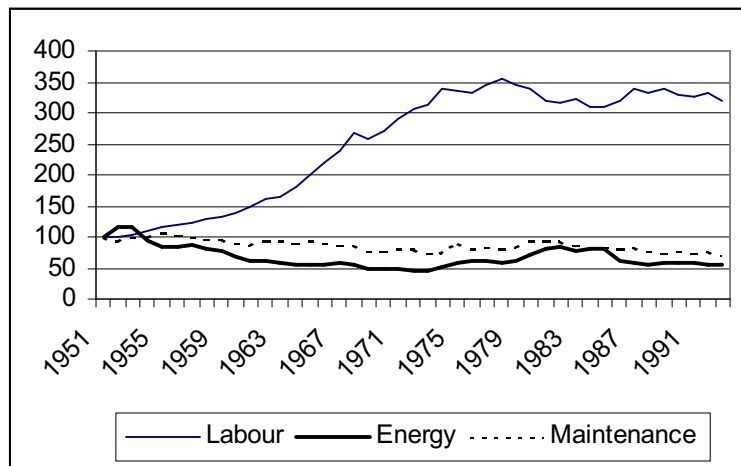


Figure 5 Factor prices over time; base NS, 1951=100

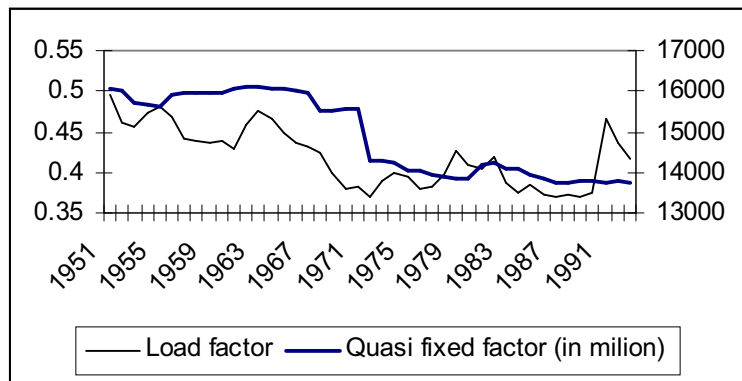


Figure 6 Load factor and capital value over time, NS

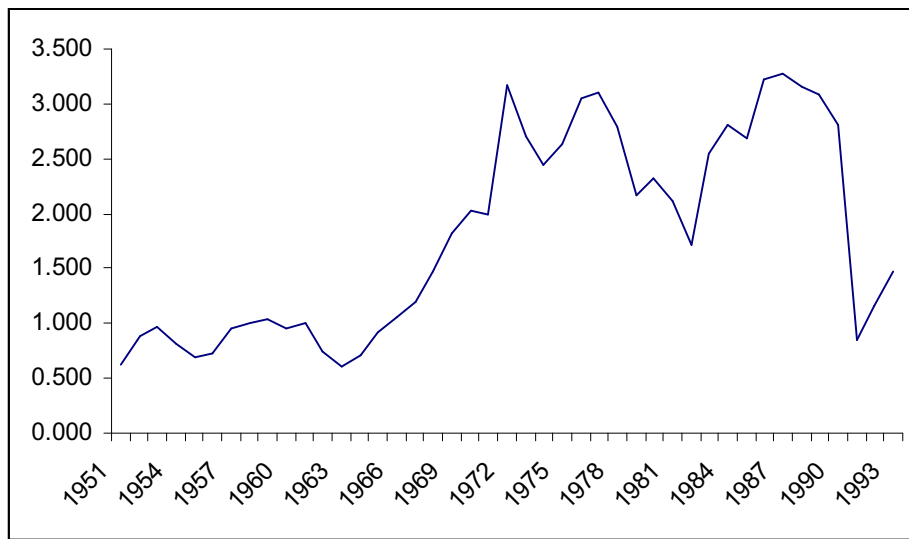


Figure 7 Evolution of returns to density